



## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

This full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/15584>

Please be advised that this information was generated on 2014-11-11 and may be subject to change.



## The Perceptual Integrity of Syllabic Onsets

ANNE CUTLER, SALLY BUTTERFIELD, AND JOHN N. WILLIAMS

*MRC Applied Psychology Unit, Cambridge, United Kingdom*

Detection time for word-initial phonemes is faster when the phoneme is followed by a vowel (as in *band*) than when the phoneme is part of a cluster (as in *brand* or *bland*). We investigated this effect as a function of the *model* which subjects were given for the target phoneme (e.g., "/b/ as in blue"). Subjects listened for word-initial target phonemes in continuous utterances. The targets occurred in single-phoneme, /r/ cluster or /l/ cluster onsets. Each type of word produced faster responses when the target had been modeled with the same onset. No such effects were found with consonant-vowel onsets; the "matching model" effect is specific to syllabic onsets, and suggests that word-initial clusters are perceived as integral units. © 1987 Academic Press, Inc.

What is the internal structure of a word like *grist*? It has five phonemes, and it has one syllable. The five phonemes are not indistinguishable equal contributors to the word's internal composition; they are grouped into the constituents of syllable structure. The vowel is called the *nucleus* of the syllable, and, as the name suggests, it is the most important component, in that it is the only one which cannot be dispensed with. What precedes the vowel is called the *onset*. The vowel and what follows it are grouped by many linguists (e.g., Fudge, 1969; Halle & Vergnaud, 1980) into a constituent called the *rime*, which in its turn has the constituents of a *peak* (the vowel) and a *coda* (what follows it); other linguists (e.g., Clements & Keyser, 1983) consider the syllable to have

only the three equal internal components of onset, nucleus, and coda.

There is no linguistic controversy about the status of the onset; it is whatever precedes the vocalic nucleus. In English, syllabic onsets can be null, or they can consist of one, two, or three consonants (*ice*, *lice*, *slice*, *splice*). Psychological evidence suggests that in speech production, syllabic onsets function as units. MacKay (1972) analyzed spontaneous speech errors and found that errors in which onsets exchanged with onsets, or in which blends of two words preserved the onset of one of the source words intact, were far more frequent than exchanges or blends in which cluster onsets were divided. Claxton (1974) asked subjects to produce words beginning with specified single consonants, consonant pairs, or consonant-vowel sequences; the mean production time for the single consonant and cluster onsets was not significantly different, but the onsets led to significantly faster word productions than onset-vowel sequences. Treiman (1983) found that novel word games which preserved syllabic onsets were easier to learn than games which divided cluster onsets (in fact, evidence from actual word games

This research was supported by a grant from the Alvey Directorate to Cambridge University, the Medical Research Council, and Standard Telecommunications Laboratories. We thank Bill Barry for technical assistance, Ian Nimmo-Smith for statistical advice, and Dennis Norris and Mary Smith for helpful discussion. Reprint requests should be addressed to the first author at MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge, CB2 2EF, United Kingdom. John N. Williams is now at the Istituto di psicologia, University of Padua, Italy.



such as pig latin shows that these, too, usually preserve onsets—Hockett, 1967).

Whether syllabic onsets can also function as units at some level of the process of speech perception is less clear. Some indirect evidence is provided by the frequently reported finding that children have difficulty in segmenting initial clusters into their constituent phonemes (Barton, Miller, & Macken, 1980; Kornfeld, 1978; Treiman, 1980). But children also have difficulty in segmenting syllables into phonemes (Liberman, Shankweiler, Fischer & Carter, 1974), whereas literate adults do not; thus speech perception data from preliterate children are not necessarily relevant to questions of speech perception in literate adults. The most relevant evidence from adult speech perception is provided by the phoneme-monitoring task, in which listeners' detection time for (usually word-initial) phoneme targets is measured. Studies using this task have repeatedly found that detection time for word-initial phoneme targets is longer if the phoneme is part of a syllable-initial cluster than if it constitutes a single-phoneme onset. Typically, it takes longer to detect /b/ in *brand* or *bland* than in *band*, and longer to detect /s/ in *street* or *sweet* than in *seat*. This finding has been reported by, among others, Foss and Gernsbacher (1983) and Cutler, Mehler, Norris, and Segui (1987), and it has been explicitly investigated by Treiman, Salasoo, Slowiaczek, and Pisoni (1982). Treiman et al. found that the inhibition of phoneme detection when the phoneme occurred in a cluster held across a range of fricatives, voiced stops, and voiceless stops. They argued that "sequences of syllable-initial consonants form a coherent perceptual unit, the syllable onset" (p. 77), and that the lengthened detection times for phonemes in clusters were caused by the need for further segmentation of the onset into its constituent phonemes.

However, there is an alternative explanation of this consistent finding. Consonants could be in some sense easier to detach

from following vowels than they are from following consonants. This suggestion is an extension of a hypothesis proposed by Cutler et al. (1986). Cutler et al. found that detection times for word-initial syllables were, for English listeners at least, consistently faster in words beginning CVCV than in words beginning CVCC; they suggested that this result might simply reflect the implementation of a phoneme-by-phoneme segmentation strategy combined with intrinsically easier segmentability of alternating sequences of consonants and vowels. Extension of this argument to syllabic onset sequences suggests that faster detection time for prevocalic than for preconsonantal consonant targets arises from intrinsic advantages of acoustic distinctiveness; by implication, then, the detection time difference does not speak to the question of whether the syllabic onset functions as a coherent unit.

It is difficult to construct a direct empirical test which will distinguish between these explanations. The acoustic distinctiveness explanation predicts that there will never be a response time advantage for cluster onsets over single-phoneme onsets; so demonstration of a response time advantage for cluster onsets under appropriate conditions should favor the unitary onset hypothesis over its rivals. However, one cannot simply measure detection time for cluster onsets and compare this with detection time for single-phoneme onsets, because target size itself exercises a substantial effect on monitoring latency (McNeill & Lindig, 1973). A slightly less direct test, though, suggests itself via the observation that phoneme-monitoring targets are customarily specified for subjects in the form of a word model. Moreover, these models have usually had single-phoneme onsets. For instance, a typical target specification might be "/b/ as in boy." Suppose that it were possible to set up expectations about syllabic onsets by manipulation of this model. If such is the case, subjects in previous experiments



have perhaps been led to expect single-phoneme onsets. However, if subjects were led instead to expect a cluster onset, they might actually detect targets faster in a cluster onset than in a single-phoneme onset. That is, targets on words like *band* would be responded to faster when the /b/ target is modeled with a single-phoneme onset than when it is modeled with any cluster onset; but targets on words like *bland* would be responded to faster when the /b/ target is modeled with a /bl/ onset rather than either a /b/ or a /br/ onset; and targets on words like *brand* would be responded to faster when the /b/ target is modeled with a /br/ onset rather than either a /b/ or a /bl/ onset. This result would suggest that in previous experiments, the single-phoneme advantage may have been due to the models that were used. Nevertheless, the existence of a matching effect would in itself be consistent with the hypothesis that syllabic onsets can function as integral units. However, it is not the result that would be predicted by the hypothesis that consonant-vowel sequences are intrinsically easier to segment than consonant clusters. This latter hypothesis holds that phoneme detection advantages with single-phoneme onsets are due to purely acoustic factors; these are constant across conditions, and should be unaffected by the model which is provided for the phoneme target. This hypothesis clearly predicts persistence of the single-phoneme advantage—responses to targets on words like *band* should always be faster than responses to targets on words like *bland* or *brand*.

In Experiment 1, below, we presented listeners with standard phoneme-monitoring materials, in which some of the target-bearing words began with single-phoneme onsets while others began with clusters. The design kept variation across conditions to a minimum. All listeners heard exactly the same sentence materials, and all performed the same phoneme detection task on each sentence. The only

difference between groups of subjects was in the models which were presented for the various phoneme targets. Each group had one of the three phoneme targets modeled with a single-phoneme onset and the other two targets modeled with cluster onsets.

The onset integrity hypothesis predicts a “matching model” effect; the intrinsic segmentability hypothesis predicts no such effect, but an advantage for single-phoneme onsets. A matching model effect and no single-phoneme advantage would therefore allow us to reject the intrinsic segmentability explanation, whereas no matching model effect would allow us to reject the onset integrity explanation.

## EXPERIMENT 1

### Materials

The target phonemes used were /p/, /b/, and /g/. They were chosen from the class of stop consonants because each can occur at the beginning of English words followed (1) by a vowel, i.e., as a single-phoneme onset; (2) by /l/ to form a Cl consonant cluster onset; and (3) by /r/ to form a Cr consonant cluster onset.

Fifty-four monosyllabic words were chosen, of which 18 began with each of the three target phonemes. Within each set of 18, six began with a consonant-vowel onset, six with a cluster of which the second member was /r/, and six with a cluster of which the second member was /l/. Within each set of six, two were nouns, two verbs and two adjectives.

A sentence was constructed for each of the 54 target words. The words occurred in varying positions in their respective sentences, but were always preceded by at least 4 words and followed by at least 1 word.

An additional 18 filler sentences were included to increase the variability of the items: 10 contained no match for the target specified, and 8 included a target-bearing word within the first 3 words of the sentence, or as the final word of the sentence.



Three alternative models were constructed for each of the three stop consonant targets. One model for each target was a word with a single-consonant onset, one was a word beginning with an /r/ cluster, and the third a word beginning with an /l/ cluster. Each set of three models was semantically coherent, in order to avoid drawing particular attention to the models. Thus one subject group heard "/p/ as in pink," "/b/ as in blue," and "/g/ as in green"; a second heard "/p/ as in proud," "/b/ as in brave," and "/g/ as in good"; and the third heard "/p/ as in plate," "/b/ as in bowl," and "/g/ as in glass." Type of onset in the models was, therefore, not perfectly counterbalanced across subject groups, but perfect counterbalancing proved impossible to achieve if each subject group was to have a semantically coherent set of reasonably high-frequency models.

One recording order was prepared for the sentences, with test and filler sentences in pseudo-random order. No successive sentences appeared to be related. The sentences were read at a normal rate by a male native speaker of British English. The nine target models were also recorded, as was a short set of practice sentences.

All recorded materials were digitized, and three tapes constructed, each with the same sentences in the same order; the tapes differed only with respect to which set of target models occurred. Each sentence was preceded by its appropriate target specification. Using digital waveform editing techniques, a timing mark on the second channel of the tape was precisely aligned with the acoustic beginning of each target word.

### *Subjects*

Subjects were 70 members of the Applied Psychology Unit panel of volunteers from the Cambridge area, aged 18–50 years, with good hearing, who were paid an honorarium for participating. Twenty-one subjects were tested in each tape condition. Data from 6 subjects who scored 60% or

below on the comprehension test, and from 1 subject who failed to take the comprehension test, were not analyzed.

### *Procedure*

The materials were presented to each subject binaurally over headphones. Written instructions asked subjects to listen to the sentences as they would in a conversation and told them that there would be a comprehension test at the end. They were instructed to "listen for a word . . . beginning with a certain sound" according to the target specified in the model shortly before each sentence, and to press the response key as soon as they heard the target sound. They were warned that not all sentences contained the target sound. At the beginning of each target word the timing signal on the other channel of the tape, inaudible to the subject, started the clock of a PDP11/23 computer. Timing stopped when the subject pressed the response key.

A yes–no recognition test of 20 sentences was administered at the end of the experiment. Half of the test items had actually been presented; the remainder were constructed from fragments of sentences which had occurred.

### *Results*

The mean score on the recognition test was 79.5%, from which we concluded that the subjects had indeed attended to sentence content.

Responses shorter than 150 ms or longer than 1500 ms were discarded (data omitted for this reason, or because the subject failed to respond, comprised 3.5% of all responses).

Separate analyses of variance were conducted with subjects and with items as random factor. The crucial comparison is the interaction of target word onset with model onset. This interaction cannot be examined directly in the subjects analysis, because of the imperfect counterbalancing



of model onset across tapes, but it can be examined directly in the items analysis.

The relevant means from the items analysis are shown in Table 1. As can be seen, the interaction of the effects of target word onset and model onset is precisely as predicted by the onset integrity hypothesis. For each target word onset, the fastest response times are achieved when the model has the same onset. The interaction of target word onset and model onset was significant ( $F_2(4,45) = 5.56, p < .001$ ). Newman-Keuls analyses showed that for each target word onset the matching model produced faster responses than each non-matching model (significant at the .05 level), except that for Cr onsets the 9-ms advantage for Cr over Cl models was not significant.

Neither the main effect of target word onset ( $F_2(2,45) = 1.11$ ) nor the main effect of model onset ( $F < 1$ ) was significant. The only other effect to reach significance was the main effect of target phoneme ( $F_2(2,45) = 5.4, p < .01$ ). Its source was that responses to /g/ were slower than responses to either /b/ or /p/.

In the subjects analysis, the main effect of subject groups was not significant, and it did not interact with the effect of target word onset or the effect of target phoneme. The main effect of target phoneme was again significant ( $F_1(2,120) = 43.75, p < .001$ ) as was, in this analysis, the main effect of target word onset ( $F_1(2,120) = 10.15, p < .001$ ). Its source was that words beginning with Cr clusters were responded

TABLE 1  
MEAN RESPONSE TIMES (MILLISECONDS) IN  
EXPERIMENT 1 AS A FUNCTION OF TARGET WORD  
ONSET AND MODEL ONSET

Target word onset	Model onset			$\bar{X}$
	C	Cr	Cl	
C	350	371	372	365
Cr	394	373	382	383
Cl	371	373	347	364
$\bar{X}$	372	373	367	

TABLE 2  
MISSED RESPONSES IN EXPERIMENT 1 AS A  
FUNCTION OF TARGET WORD ONSET AND  
MODEL ONSET

Target word onset	Model onset		
	C	Cr	Cl
C	8	15	16
Cr	6	5	13
Cl	20	19	14

Note. Number of data points per cell = 378.

to slower (383 ms) than the other two word types (365 and 364 ms, respectively). The crucial comparison can be assessed in this analysis via the three-way interaction of subject groups with target word onset and target phoneme. This was significant ( $F_1(8,240) = 2.55, p < .01$ ). For each phoneme and each onset type, the fastest responses should have been produced by that subject group which had the matching model onset. This was indeed the case in 16 of the 18 such comparisons (for single-phoneme onsets with /g/, subjects who had the model “good” were 5 ms slower than subjects who had the model “green,” and for Cr onsets with /p/, subjects who had the model “proud” were 3 ms slower than subjects who had the model “plate”). The mean advantage for matching model over nonmatching model was 21 ms (24 ms across the 16 positive comparisons).

Finally, this matching versus non-matching model effect can be tested directly if we ignore the effects of target phoneme and of model word onset and simply collapse across the matching and nonmatching responses, respectively. The 21-ms difference is significant in both such analyses ( $F_1(1,60) = 29.51, p < .001$ ;  $F_2(1,45) = 19.78, p < .001$ ; min  $F'(1,94) = 11.84, p < .001$ ).

Subjects’ errors (missed responses) were also inspected. As can be seen from Table 2, although there were few errors, the data were in line with the reaction time data. For each target word onset, the fewest



errors were made when the model had the same onset.

### *Discussion*

As predicted by the onset integrity hypothesis, Experiment 1 showed a significant "matching model" effect. Contrary to the prediction of the intrinsic segmentability hypothesis, there was no sign of an overall advantage for words with single-phoneme onsets. Targets on words with single-phoneme onsets were responded to fastest when the target had been modeled with a single-phoneme onset; targets on words beginning with Cr were responded to fastest when the target had been modeled with Cr; and targets on words beginning with Cl were responded to fastest when the target had been modeled with Cl. Although all subjects heard the same set of materials, and all performed the same single-phoneme detection task, the model apparently set up expectations about syllabic onsets and led to faster responses when the target-bearing words confirmed these expectations than when they did not.

We are therefore in a position to reject the simple acoustically based explanation of phoneme detection time advantages in single-phoneme onsets. In this experiment there was no overall single-phoneme onset advantage of this kind. In contrast, single-phoneme onsets were only associated with faster responses when the target had been modeled with a single-phoneme onset. This suggests that, at least for stop consonants, there is no intrinsic difference in segmentability for initial consonants followed by vowels versus initial consonants followed by consonants (in this case, glides).

Of course, the acoustic-phonetic characteristics of stops do differ in these two types of environment. However, the differences cannot support the three-way distinction between C, Cr, and Cl onsets reflected in our subjects' responses. Take, for instance, voice onset time variation. Klatt (1975) computed average VOT values for stops in single-phoneme and cluster onsets.

For [p], the mean values he found were 47 ms in C onsets, 59 ms in Cr onsets, and 61 ms in Cl onsets; for [b], 11, 14, and 13 ms, respectively; and for [g], 27, 35, and 26 ms, respectively. Thus for [p], VOT in Cr and Cl onsets is virtually identical, although both differ from C onsets. The mean response times of Experiment 1 show a quite different pattern: mean response times for [p] words were almost identical for C and Cl onsets (345 and 347 ms, respectively), while both differed from Cr onsets (384 ms). For [b], mean response times to C onsets (339 ms) and Cl onsets (343 ms) were alike, and different from mean response time to Cr onsets (382 msec)—again, completely unlike the VOT pattern. For [g], mean response times did not differ greatly between the three onsets (C: 409 ms; Cr: 384 ms; Cl: 402 ms), whereas the VOT values varied much more than they did for the other voiced stop in question, [b]. We would argue that the matching model effect is not driven by acoustic-phonetic similarity, but instead reflects subjects' expectations at a more abstract level.

However, while the matching model effect suggests that the models influenced subjects' expectations, the expectations may not have been specifically about syllable structure. That is, the matching model effect may be just as easy to invoke with, say, onset-plus-nucleus combinations. There is much evidence of various matching effects in monitoring tasks. For instance, McNeill and Lindig (1973) found that response times were facilitated when the type of target (e.g., phoneme, syllable, word) matched the type of stimulus material—words were detected fastest in lists of words, syllables in lists of syllables, and so on. Healy and Cutting (1976) found that a match between the level of the target and the level of the response item facilitated detection even when the stimulus lists were heterogeneous. Mills (1980a) found that acoustic identity of target and response item speeded syllable-monitoring responses.



Four experiments specifically investigated the role of the following vowel in the processing of initial consonants. Wood and Day (1975) found that subjects who were asked to classify CV stimuli on the basis of either the consonant or the vowel were unable to ignore variation in the dimension irrelevant to the classification task. Shand (1976) showed that variation in the second syllable of CVCV stimuli had a similar effect. Swinney and Prather (1980) found that subjects monitoring for the target /b/ in lists of nonsense syllables responded faster when the vocalic nuclei of the syllables were relatively homogeneous than when they were very variable. This effect held even when the subjects were explicitly told that several different vowels would occur in the stimulus items. Finally, in the nearest analog to our target model effect, Mills (1980b) investigated phoneme monitoring in lists of syllables as a function of the target specification, which was also a CV syllable. Although subjects were instructed to ignore the vowel in the target specification, they apparently could not; when the target was given as /be/, for example, subjects were faster detecting /b/ in /be/ than in /bo/.

These results suggest that subjects in our Experiment 1 may have constructed specific expectations about the acoustic form of the response items rather than the abstract structural expectations which the onset integrity hypothesis implies. However, results with lists of short, isolated syllables are not directly relevant to the processing of continuous speech. In our experiment subjects had to segment the signal in order to ensure that the target phoneme been detected in the correct word-initial position. By stipulating a minimum level of performance on the subsequent recognition test, we ensured that subjects were processing the content of the stimuli as they were listening for the phoneme target. Thus our subjects undertook a much more complex task than Mills' or Swinney and Prather's subjects, and they were presum-

ably less likely to be able to retain an acoustic representation of the target specification than subjects in the simpler experiments.

In order to establish whether or not the matching model effect on phoneme detection in continuous speech is specific to syllabic onsets, we next investigated the effect of following context on detection of stops in single-phoneme onsets; the experiment was thus similar to Mills' study with isolated syllables. As in the preceding experiment, each phoneme target had alternative models; but the models in this case each had the same onset. All the rest of the model word differed. For instance, the two models for /b/ were "/b/ as in Ben" and "/b/ as in Billy." The response items varied in the degree to which they overlapped with the model—in the initial (target) phoneme only, in the initial consonant and its following vowel, or in the whole initial (CVC) syllable.

## EXPERIMENT 2

### Materials

Four target phonemes were chosen for this experiment: the three used in Experiment 1 plus /t/. Two target models were chosen for each phoneme; these were all proper names: Pam and Polly, Ben and Billy, Gus and Gary, Tim and Terry. Eight words beginning with each phoneme were chosen. Of these, two began with the same three phonemes as each of the models, while a further two began with the same two phonemes as each of the models. For example for /t/ the words were *timid*, *timber*, *terror*, *terrace*, *tickle*, *tissue*, *tepid*, and *testing*. Thus a subject who heard "/t/ as in Tim" would respond to two target-bearing words matching the model on three phonemes (*timid* and *timber*), two matching it on two phonemes (*tickle* and *tissue*), and four in which only the target phoneme itself matched, i.e., the match to the model was one phoneme (*terror*, *terrace*, *tepid*, and *testing*). A sentence was



constructed for each word. As in Experiment 1, the response item was always preceded in its sentence by at least four words and followed by at least one word.

The sentences, a set of practice sentences, and the target specifications were recorded and digitized as in Experiment 1. Two tapes were prepared, each containing all 32 experimental sentences, the 8 practice sentences and 16 filler sentences. (Of these, 8 had no occurrence of the specified target, and 8 had targets occurring within the first three words of the sentence, or sentence-finally.) For one tape, the target models were “/p/ as in Polly”; “/b/ as in Ben”; “/g/ as in Gary”; and “/t/ as in Tim.” The other tape used the other four models. Target specifications and timing signals were placed as in Experiment 1.

### Subjects

Fifty-four members of the same subject population as in Experiment 1 took part and were paid for participating. The data from 4 subjects who scored 60% or below on the comprehension test were discarded. Of the remaining subjects, 25 heard each tape.

### Procedure

The procedure was as in Experiment 1. Testing was carried out by the same experimenters and in the same room with the same equipment.

### Results

The mean recognition test score was 81.8%, indicating that again subjects were attending to sentence content.

Responses shorter than 150 ms or longer than 1500 ms were discarded (data omitted for this reason, or because the subject failed to respond, comprised 1.8% of all responses).

The items were divided into two groups—those which matched one or other model on three phonemes, and those which matched one or other model on two phonemes. Thus for the /t/ words, *timid*,

*timber*, *terror*, and *terrace* were classified as three-phoneme matches (with “Tim” and “Terry,” respectively), while *tickle*, *tissue*, *tepid*, and *testing* were two-phoneme matches (with the same models). Subjects who had heard the model “Tim,” for example, thus had heard two words from the first group which fully matched the model while the other two matched it only on the initial phoneme, and two words from the second group which matched the model on two phonemes, while the other two words matched it only on the initial phoneme. The same was true of subjects who had heard the model “Terry.” This division therefore allowed perfectly counter-balanced analyses to be conducted across both subjects and items, with target type (three-phoneme match versus two-phoneme match) and matching level (high versus low, where low was a match on initial phoneme only) as two-level factors.

The means for the interaction of these two factors are given in Table 3. It can be seen that when the model matched the target word on three phonemes, responses took on average 382 ms, when it matched on two phonemes 381 ms, and when it matched on only one phoneme 382 ms (the latter number is obtained by collapsing across the “low” column, in which both cells represent a match of the target phoneme only).

The only effect to reach significance in either analysis was the effect of target phoneme; again, responses to words beginning with /g/ were longest. Responses to words beginning with /t/ or /p/ were intermediate, and responses to words beginning

TABLE 3  
MEAN RESPONSE TIMES (MILLISECONDS) IN  
EXPERIMENT 2 AS A FUNCTION OF TARGET TYPE  
VERSUS MATCH TO MODEL

Target type	Level of match to model	
	High	Low
Three phoneme	382	380
Two phoneme	381	384



with /b/ were fastest ( $F_1(3,144) = 12.47$ ,  $p < .001$ ;  $F_2(3,24) = 3.26$ ,  $p < .05$ ). No other main effect or interaction reached significance.

The very small set of missing data was also inspected. Where the model matched the target word on three phonemes, subjects missed the target on 6 occasions (of 400, i.e., 1.5%); where the model matched the target on two phonemes, the target was missed on 8 occasions (2%); where the model matched the target on only the initial phoneme, the target was missed on 15 occasions (of 800, i.e., 1.87%).

### Discussion

The high-level matches between model and response word produced no faster responses than the low-level matches. We have therefore failed to replicate the matching model effect found in Experiment 1 with syllabic onsets. Thus the matching model effect must not be the same as the acoustic match effects demonstrated by Mills (1980b) and Swinney and Prather (1980) in lists of isolated and acoustically simple syllables.

The repeated finding that responses to [g] targets were longer than responses to [b], [p], or [t] is, nevertheless, likely to reflect acoustic factors. Second- and third-formant transitions are more variable across following contexts in velar stops than in bilabial or alveolar stops (Fant, 1969; Gay, 1979; Lehiste & Peterson, 1961). Note that if there were an acoustic match effect operating in Experiment 2, it might be expected to be most obvious where acoustic variability in the realization of a given phoneme is greatest; that is, [g] words might be more likely to show an effect than [p], [b], or [t] words. In fact, [g] words showed the *least* trace of a matching model effect, with mean response times of 404 ms when the model matched, 402 ms when it did not.

We claim that the matching model effect operates at a more abstract level than previous acoustically based effects. Experi-

ment 2 suggests that at this more abstract level it does not matter what vowel follows the target. Variation in quality of the following vowel makes a difference at the acoustic level, but not at a phonemic level. The results of Experiment 1, however, suggest that it does matter at this level whether an initial phoneme is part of a cluster or constitutes a syllabic onset by itself. That is, syllable structure is relevant at this relatively abstract level of representation. The matching model effect in the specification of phoneme targets is specific to syllabic onsets.

Since this argument rests upon the presence of an effect in Experiment 1 but the absence of an effect in Experiment 2, however, it is important to compare the two experiments. First, note that the power of Experiment 2 is no weaker than that of Experiment 1 (400 individual responses contribute to each cell of Table 3, compared with 378 per cell in Table 1). Second, response times in Experiment 1 (overall mean 367 ms), in which there was a significant effect, were not slower than those in Experiment 2 (overall mean 382 ms), in which there was no effect. Third, we can statistically compare the magnitude of the effect in the two experiments. In Experiment 1, collapsing across the cases in which the target did *not* match the model (e.g., responses to C-onset and Cr-onset words when the subject had heard a Cl-onset target) produced a two-level factor of matching versus nonmatching model. Mean response time when the target did match the model was 356 ms, mean response time when the target did not match the model was 377 ms. (These numbers are the mean of the three matching cells on the diagonal of Table 1 and the mean of the remaining six cells of Table 1, respectively.) This difference can be compared with that found in Experiment 2, in which mean response time when the target did match the model was 381 ms and mean response time when the target did not match the model was 382 ms. (These numbers are the column means



of Table 3.) A test for homogeneity of variance showed that the error variances in the two experiments were not significantly different ( $F(24,45) = 1.65$ ). This allowed the computation of a  $t$  statistic derived from the size of each matching effect taking into account its standard error. This was significant ( $t(69) = 2.52, p < .02$ ), indicating that the magnitude of the difference between matching and nonmatching models was reliably larger in Experiment 1 than in Experiment 2. An analysis of variance with experiments and the two-level matching factor as the only independent variables similarly showed that the interaction between these two variables was significant ( $F_2(1,84) = 6.74, p < .015$ ).

### GENERAL DISCUSSION

The phoneme-monitoring task is aptly named. Successful detection of a sound such as /b/ in words as diverse as *barn*, *beat*, *brevity*, *bludgeon*, and *balloon* requires a relatively abstract representation of the target, one which will be impervious to variations in closure duration, burst amplitude, and formant transitions which may render the acoustic form representing /b/ in one word quite different from the acoustic form of /b/ in another word. Experimenters using phoneme monitoring have long known this, and have presented target specifications in a constant form despite considerable variation across response items. Moreover, the target specification typically emphasizes the need to abstract away from acoustic form by first *specifying* the target in a relatively neutral form, i.e., followed by a schwa, and then *modeling* it on a particular word. Indeed, although phoneme targets are customarily spoken with a schwa before being modeled on a word, response items in which the target is followed by a schwa—i.e., response items beginning with unstressed syllables—elicit slower responses than those in which the target is followed by a full vowel. This effect of stress is robust and has been found in many experiments. An examination of

item means for the sentences used by Cutler and Foss (1977) shows that targets followed by schwa were detected on average 89 ms slower than other targets, both when the target-bearing word was accented and when it was not accented.

Mills' (1980b) finding suggests, it is true, that when response lists are short and composed solely of CV syllables, subjects may retain an acoustic representation of the target specification. The results of our second experiment, however, show that subjects do not retain acoustic impressions while listening to continuous meaningful speech.

There is, of course, no reason why they should. The task of speech understanding does not require retention of acoustic form. In contrast, efficient understanding puts a premium on rapid conversion of acoustic form into higher level representations. Listeners must learn to abstract away from irrelevant acoustic variations due to individual speaker characteristics, and seek instead the more abstract underlying forms.

Moreover, such abstract representations, not least among them a representation of syllable structure, are often required in everyday language processing. Production or recognition of rhyme depends on an appreciation of syllable structure—two words rhyme when they share all segments which follow the onset of the stressed syllable. Rhyme is not the only poetic form dependent upon syllable structure—medieval poetry made extensive use of assonance (in which syllabic nuclei are matched) and alliteration (in which onsets are matched). Alliteration as a poetic form may now be called upon principally by creators of brand names and advertising slogans; but its effectiveness in such uses presumably rests upon listeners' ability to notice and appreciate it, or to use it in recall. Finally, language games such as pig latin also demonstrate exploitation of our knowledge of syllable structure.

This level of representation is apparently not susceptible to simple form priming.



Slowiaczek and Pisoni (1986) failed to find facilitation of auditory lexical decision responses when the target word was preceded by another word which shared one or more phonemes with the target. (Although Slowiaczek & Pisoni's prime-target pairs sometimes shared onset, and sometimes onset and nucleus, the authors do not report these two types of overlap separately.) Compare these results with the facilitation found by Slowiaczek, Nusbaum, and Pisoni (1987), who required subjects to identify words in noise. Under these conditions, words were more accurately reported when they were preceded by other words having one or more phonemes in common with them. That is, when acoustic information is exiguous, hints help. When acoustic information is adequate, however, words are rapidly translated into a more abstract representation without reference to similar decisions about neighboring words.

This, too, makes sense in the context of normal speech recognition. Acoustic forms are highly context dependent: for example, a given stretch of speech might represent /p/ before one vowel, /k/ before another. Thus direct priming of the translation from acoustic form to higher level representation might be rather risky, in that it could often mislead. On the other hand, when the acoustic information is insufficient for certain identification, recent translation decisions could set up a response bias.

This brings us, finally, to consideration of the precise mechanism by which the matching model effect operates. We see three possibilities. First, the abstract representation of the phonemic target could itself be affected by the model, so that "/b/ as in bowl" leads to the formation of a different target representation from, say, "/b/ as in blue." This suggestion seems to us incompatible with the low error rate in Experiment 1 (Table 2 shows that the highest error rate in any condition was 5.3%), since if the target representation only imperfectly matches two-thirds of the response items, a

high error rate should be expected. Second, the matching model effect might be due to priming not of the translation from acoustic form, but of the assignment of syllabic structure. Third, the effect might be due to postdecision response bias of some kind. Our results do not allow us to distinguish unequivocally between these latter possibilities. However, if the assignment of syllabic structure can indeed be primed, it should be possible to demonstrate such priming in any task in which responses require a representation of syllable structure, such as, for instance, the language games studied by Treiman (1983).

What is clear is that the effect is not a simple acoustic match, but takes place at an abstract level at which syllabic structure is also relevant. It is also clear that the previously found phoneme-monitoring response advantage for targets in single-phoneme onsets is not due, as suggested by Treiman et al. (1982), to the extra time necessary for subdivision of a cluster onset into its constituent phonemes; nor is it due to intrinsically easier perception of consonants in vocalic as opposed to consonantal environments, since in either case it should have appeared irrespective of model. The fact that the models do exercise differential effects can only be explained in terms of the perceptual integrity of syllabic onsets at some level of language processing; and it certainly provides further evidence in favor of the claim put forward by Treiman and her colleagues, that syllabic structure is psychologically real.

## APPENDIX

### *Target-Bearing Words and Example Sentences*

#### Experiment 1

back, bike, boast, bind, bold, best, bridge, brand, break, bring, broad, brash, blend, blot, bleed, blow, bleak, blank; games, goose, gain, gulping, gallant, gold, grid, gross, greeted, grasp, grave, grand, glut, glee, glance, glazing, glib, glad; pork, post, pawn, perching, pale, pert, prize, prank, print, prove,



prim, prone, pledge, plight, pleased, plunge,  
plumper, plain

#### Example sentences

The young chess champion showed no tendency to *boast* about his achievements.

The local firm's estimate for the *glazing* job was beaten by a London company.

It was announced at the fair that a special *prize* would be awarded for the most well-stocked stall.

#### Experiment 2

builder, billiards, bitter, biscuit, benches, bending,  
bury, beggar; garage, garret, gather, gallon, gusty,  
gusto, governor, gunshot; pamphlet, pamper,  
packing, passage, pollen, policy, poverty, posture;  
timid, timber, tickle, tissue, terror, terrace, testing,  
tepid.

#### Example sentences

The demonstrators tried to give a *pamphlet* to each worker entering the factory.

When the salesman had nearly finished *packing* his bags, he realised that he'd forgotten his shaving kit.

Despite his doctor's opinion, the man doubted that his *posture* caused his backache.

### REFERENCES

- BARTON, D., MILLER, R., & MACKEN, M. A. (1980). Do children treat clusters as one unit or two? *Papers and Reports on Child Language Development*, 18, 93-137.
- CLAXTON, G. L. (1974). Initial consonant groups function as units in word production. *Language and Speech*, 17, 271-277.
- CLEMENTS, G. N., & KEYSER, S. J. (1983). *CV Phonology*. Cambridge, MA: MIT Press.
- CUTLER, A., & FOSS, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, 20, 1-10.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, 19, 341-377.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385-400.
- FANT, G. (1969). Stops in CV syllables. *Speech Transmission Laboratory, Stockholm: Quarterly Progress and Status Report*, 4, 1-25.
- FOSS, D. J., & GERNSBACHER, M. A. (1983). Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, 22, 609-632.
- FUDGE, E. C. (1969). Syllables. *Journal of Linguistics*, 5, 253-286.
- GAY, T. (1979). Coarticulation in some consonant-vowel and consonant cluster-vowel syllables. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication*. New York: Academic Press.
- HALLE, M., & VERGNAUD, J. R. (1980). Three dimensional phonology. *Journal of Linguistic Research*, 1, 83-105.
- HEALY, A., & CUTTING, J. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, 15, 73-83.
- HOCKETT, C. F. (1967). Where the tongue slips, there slip I. In *To honor Roman Jakobson*. The Hague: Mouton.
- KLATT, D. H. (1975). Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18, 686-706.
- KORNFELD, J. (1978). Implications of studying reduced consonant clusters in normal and abnormal child speech. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Language development and mother-child interaction*. New York: Plenum.
- LEHISTE, I., & PETERSON, G. E. (1961). Transitions, glides and diphthongs. *Journal of the Acoustical Society of America*, 33, 268-277.
- LIBERMAN, I. Y., SHANKWEILER, D., FISCHER, F. W., & CARTER, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18, 201-212.
- MACKAY, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3, 210-227.
- MCNEILL, D., & LINDIG, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, 12, 431-461.
- MILLS, C. B. (1980a). Effects of the match between listener expectancies and coarticulatory cues on the perception of speech. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 528-535.
- MILLS, C. B. (1980b). Effects of context on reaction time to phonemes. *Journal of Verbal Learning and Verbal Behavior*, 19, 75-83.
- SHAND, M. A. (1976). Syllabic versus segmental perception: On the inability to ignore "irrelevant" stimulus parameters. *Perception & Psychophysics*, 20, 430-432.
- SLOWIACZEK, L. M., & PISONI, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, 14, 230-237.
- SLOWIACZEK, L. M., NUSBAUM, H. C., & PISONI,



- D. B. (1987). Acoustic-phonetic priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 64-75.
- SWINNEY, D., & PRATHER, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics*, 27, 104-110.
- TREIMAN, R. (1980). *The phonemic analysis ability of preschool children*. Unpublished doctoral dissertation, University of Pennsylvania.
- TREIMAN, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, 15, 49-74.
- TREIMAN, R., SALASOO, A., SLOWIACZEK, L. M., & PISONI, D. B. (1982). Effects of syllable structure on adults' phoneme monitoring performance. *Progress Report No. 8*, Indiana University, Speech Research Laboratory.
- WOOD, C. C., & DAY, R. S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics*, 17, 346-350.
- (Received July 28, 1986)
- (Revision received March 3, 1987)